

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Machine Learning

LECCIÓN 01


















Lección 1: Preparación del entorno y carga de datos

ÍNDICE

- ✓ Introducción
- ✓ Objetivos
- ✓ Python y Machine Learning
- ✓ Preparación del entorno Python para Machine Learning
- ✓ Carga de datos
- ✓ Conclusiones

INTRODUCCIÓN

- Python y su creciente uso.
- Características de Python.
- Ranking Tiobe.
- Ciencias de los datos y Machine Learning con Python.

Oct 2021	Oct 2020	Change	Programming Language	Ratings	Change
1	3	▲	 Python	11.27%	-0.00%
2	1	▼	 C	11.16%	-5.79%
3	2	▼	 Java	10.46%	-2.11%
4	4		 C++	7.50%	+0.57%
5	5		 C#	5.26%	+1.10%
6	6		 Visual Basic	5.24%	+1.27%
7	7		 JavaScript	2.19%	+0.05%
8	10	▲	 SQL	2.17%	+0.61%
9	8	▼	 PHP	2.10%	+0.01%
10	17	▲	 Assembly language	2.06%	+0.99%
11	19	▲	 Classic Visual Basic	1.83%	+1.06%
12	14	▲	 Go	1.28%	+0.13%
13	15	▲	 MATLAB	1.20%	+0.08%
14	9	▼	 R	1.20%	-0.79%
15	12	▼	 Groovy	1.18%	-0.05%
16	13	▼	 Ruby	1.12%	-0.05%
17	16	▼	 Swift	1.11%	+0.02%

OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Conocer la potencialidad de Python y sus módulos en Machine Learning.
- 2 Instalar y configurar el ecosistema Python para trabajar en Machine Learning.
- 3 Conocer los formatos de archivo que contienen los datos en Machine Learning.
- 4 Cargar conjuntos de datos para Machine Learning en Python.

Scipy

Ecosistema de software de código abierto basado en Python para las matemáticas, la ciencia y la ingeniería.



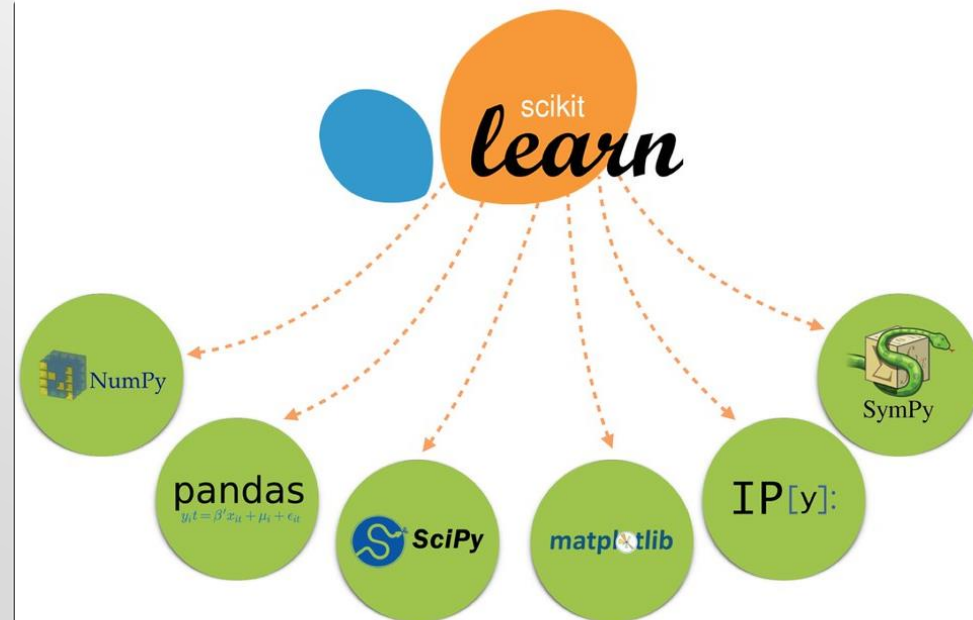
Scikit-learn

Herramientas necesarias para desarrollar y ejecutar proyectos de Machine Learning en Python.

Posibilidad de:

- ✓ aprender sobre Machine Learning
- ✓ desarrollar modelos
- ✓ ponerlos en producción

todo en el mismo marco y con el mismo código.



Instalación de Scipy

Las maneras más populares son:

- Gestión de paquetes del propio sistema operativo.
- Gestión de paquetes de Python: *pip*.
- Anaconda.

PROBAR LA INSTALACIÓN



A screenshot of the SciPy.org website. The header is blue with the SciPy.org logo. The main content area is white and titled "Installation". It lists installation methods: Distributions, pip, Package Manager, Source, and Binaries. It also mentions that these methods differ in ease of use, coverage, and maintenance. A section titled "Scientific Python Distributions (recommended)" lists Python distributions like Anaconda, Miniconda, and WinPython. The right sidebar contains links to "About SciPy", "Getting started", "Documentation", "Install", "Bug reports", "Codes of Conduct", "SciPy conferences", "Topical software", "Citing", "Cookbook", "Blogs", and "NumFOCUS". At the bottom, there is a "Table of Contents" section with links to "Installation" and "Scientific Python Distributions".

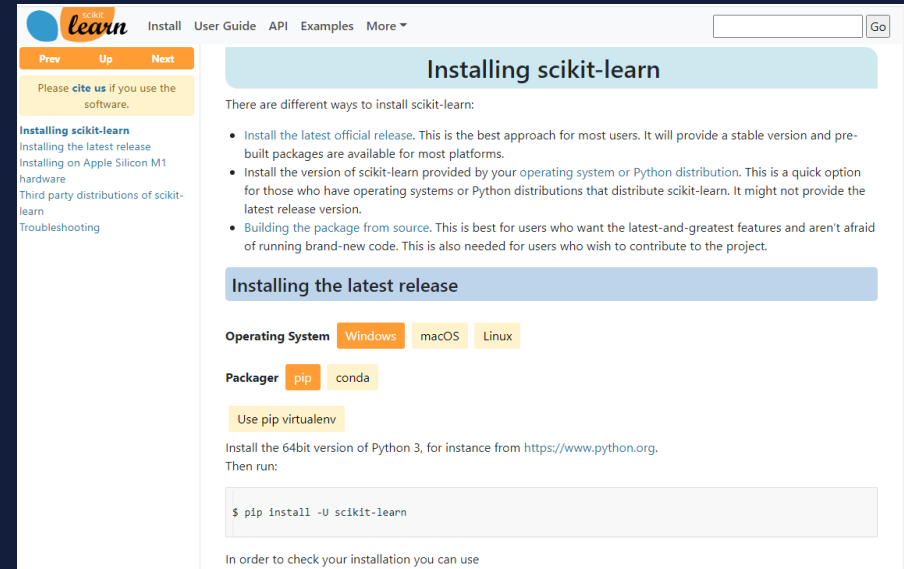
Instalación de scikit-learn

Las maneras más populares son:

- Gestión de paquetes del propio sistema operativo.
- Gestión de paquetes de Python: *pip*.
- Anaconda.

Integración es Spyder

PROBAR LA INSTALACIÓN



The screenshot shows the official scikit-learn website's installation page. It features a navigation bar with links for 'Install', 'User Guide', 'API', 'Examples', and 'More'. The main heading is 'Installing scikit-learn'. Below this, it states 'There are different ways to install scikit-learn:' and lists three methods: installing the latest official release, installing from a pre-built package for specific operating systems or Python distributions, and building from source. A section titled 'Installing the latest release' provides instructions for Windows, macOS, and Linux, and offers options for using 'pip' or 'conda' as the packager. It also includes a terminal command: `$ pip install -U scikit-learn`.



CARGA DE DATOS

CSV (Comma Separated Values)



Conjuntos de datos

- Repositorios



Conjunto de datos “Indian Liver Patient”

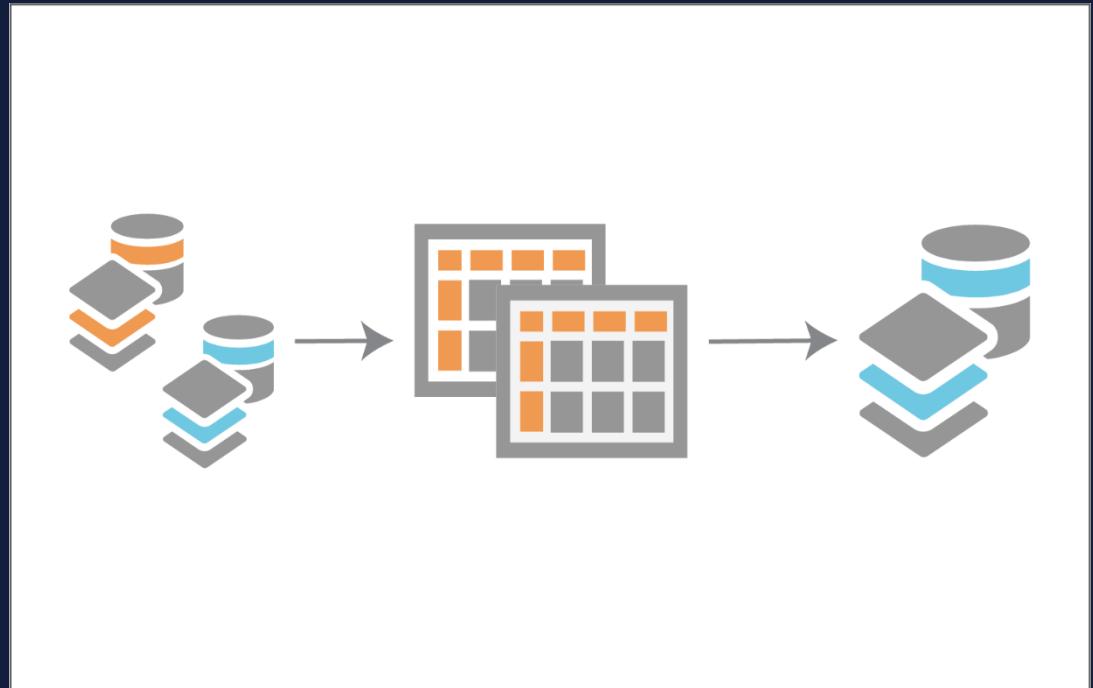
- Información sobre pacientes de hígado.
- Problema de clasificación.
- Atributos de entrada numéricos, atributos de entrada nominales
- Variable de salida a predecir es binaria.

'age'
'gen'
'tbili'
'dbili'
'alkphos'
'sgpt'
'sgot'
'tp'
'alb'
'ag'
'class'

Carga de un conjunto de datos

Varias opciones:

- a) Python nativo
- b) Numpy
- c) Pandas



MUCHAS GRACIAS POR SU ATENCIÓN



jperez@grupomainjobs.com



Javier Pérez Rodríguez
www.linkedin.com/in/perezxavi



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados